Thesis/
Reports
CSU

Mining and Water Quality:
Abandoned and Inactive Coal Mines

Final Report

Final Report for
Mining and Water Quality: Abandoned and Inactive Coal Mines

Cooperative Agreement No. 28-C3-764

between

Colorado State University, Department of Forest Sciences

and

United States Forest Service
Rocky Mountain Forest and Range Experiment Station
Research Work Unit No. 4852

INTRODUCTION

The goal of this project was to assemble a spatial database that can be used to gain a better understanding of the water quality problems posed by abandoned and inactive coal mines throughout the United States. Toward this end it was necessary to combine data about mines from two disparate sources in a manner that kept duplicate information to a minimum. Prior to this project a portion of the Minerals Availability System (MAS) non-proprietary database was acquired from the U.S. Bureau of Mines. The data acquired included the spatial locations and characteristics of mines located throughout the United States. The spatial location of each mine and a portion of the mine characteristic information had been recorded in an ARC/INFO geographic information system (GIS) format prior to the start of this project. However, since this database does not contain information for all mines in the U.S. it was necessary to acquire additional data from other sources and merge this data with the MAS database. This paper describes the process of building a spatial database of mines data acquired from the U.S. Geological Survey (USGS) and the process of identifying duplicate mine data within this database, within the MAS database, and between the two databases.

METHODS

I. BUILDING THE SPATIAL DATABASE

The Mineral Resources Data System (MRDS) database was acquired from the USGS in an ASCII formatted file. It was first necessary to convert this ASCII file into an ARC/INFO GIS file. Since complete location information is necessary when building a spatial database, each record in the MRDS database was examined for valid location information. Programs were written in the C programming language in order to evaluate the location information for each MRDS record. These programs checked each MRDS record for either valid Universal

LIBRARY COPY
ROCKY MTN. FOREST & RANGE
EXPERIMENT STATION

Transverse Mercator (UTM) coordinates or valid latitude and longitude coordinates. For those records with UTM coordinates, valid was defined as a UTM northing value less than 10,000,000, a UTM easting value less than 1,000,000, and a UTM zone value between 1 and 60. For those records with latitude and longitude coordinates, valid was defined as having at least degrees and minutes for both latitude and longitude, and a degree value less than 360, a minute value less than 60, and a second value less than 60, if such a value existed. No attempt was made to compare each record's location information (either UTM or latitude and longitude coordinates) to the state and county information. The programs were designed simply to retain records of mine locations that had logical or sufficiently detailed location information (Table 1).

MRDS Records acquired from USGS	78604
Portion with valid UTM coordinates	37633
Portion with valid lat/long coordinates	31869
Portion with invalid coordinates	9102
Total records in MRDS spatial database (i.e., those records with valid locational information)	69502
Total records in MAS spatial database	207242

Table 1. Summary of records in MRDS and MAS spatial database.

Once records with inadequate or illogical location information were eliminated the data was split into spatial and nonspatial components. A simple ASCII file of the spatial data was created and the ARC/INFO 'generate' command was used to import this spatial information into an ARC/INFO spatial database. Once in ARC/INFO the records with valid UTM values were converted into latitude/longitude coordinates and merged with the valid latitude/longitude records (Table 1).

In order to complete the process of building a spatial database from the MRDS data it was necessary to import the nonspatial data (mine characteristic information) into ARC/INFO. Because of the large number of fields for each record in the MRDS database and the large number of characters for many of the fields, it was necessary to import the data into several ARC/INFO files. C programs were written in order to divide the MRDS database into several files and to properly format the files for import. These files were then individually imported into ARC/INFO.

Problems were encountered using two of the three ARC/INFO commands available for data import. The single successful import operation involves importing fixed length fields; unfortunately, due to the large files created when using fixed length fields, only a portion of the MRDS mine characteristic data could be practically imported and linked to the ARC/INFO

spatial data. The MRDS fields selected for importing included those fields that were to be used when searching for duplicate mines (Table 2). An alternate method of importing the remaining MRDS data into ARC/INFO by transferring the data into dBASE IV formatted files and converting these files to ARC/INFO export files using PC ARC/INFO has been successfully tested with 500 records. This technique could be used to import the remaining MRDS data.

MRDS Record Number	Elevation	Latitude
Report Date	Northing	Longitude
Synonym Name	Easting	Commodity Codes
County	Zone	Development Status
State	Township	Deposit Type
Land Status	Range	Workings Type
Quad Name	Section	Update Date
Quad 1 Scale	Site Name	

Table 2. MRDS mine characteristic information imported into ARC/INFO.

As was mentioned previously, only a portion of the MAS mine characteristic information was present in an ARC/INFO spatial database when this project began. The available data included some but not all of the information that was needed when searching for duplicate mines. A program was written to extract the additional fields from the MAS database and create an ASCII file formatted for import into ARC/INFO. The fields were then imported into ARC/INFO and linked to the existing MAS ARC/INFO spatial database (Table 3).

The commodity information for the MAS database was contained in a separate ARC/INFO GIS file. Unfortunately this file was in a format that would not allow comparisons in ARC/INFO with the MRDS commodity data. This MAS commodity file contained several records (1 record for each commodity produced) for each MAS mine. In order to be able to compare the MAS commodity information to the MRDS commodity information it was necessary to reformat the MAS commodity information. The data from the MAS commodity file was exported to an ASCII file and reformatted using a C program. The resulting ASCII file contained 1 commodity record for each MAS mine, where each record could contain multiple commodities. This file was imported into a new ARC/INFO GIS file and linked to the MAS database.

MAS Sequence Number	County Code
Deposit Name	State Code
Type of Operation	Elevation
Current Status	

Table 3. MAS mine characteristic information imported into ARC/INFO.

II. EVALUATING EXISTING CROSS-REFERENCES

Once the necessary MRDS and MAS data was in the proper ARC/INFO format, the databases could be evaluated for duplicate mine information. First an attempt was made to evaluate existing cross-referencing information between databases in order to find existing known duplicates. The MAS database includes a field that contains the record number of any known corresponding record in the MRDS database. A C program was written to screen the MAS database and create an ASCII file containing the MAS sequence number and the associated MRDS record number of all MAS records with such cross-referencing information (Table 4).

MAS records with cross-reference to MRDS	19225
MRDS records with cross-reference to MAS	2260
Cross-references previously identified in MAS	87
MRDS records cross-referenced to same MAS mine	54
Records with unique cross-reference between MRDS and MAS	21344

Table 4. Summary of records with cross-reference information.

The MRDS database also contains a cross-reference field (the 'File Link Identification' field). This field is not strictly a cross-reference to the MAS database, but can contain codes that indicate a cross-reference to a number of other mines databases. A C program was written to evaluate this field in order to find records that contained the MAS cross-reference code and an MAS sequence number. For these records, an ASCII file was created that contained the MRDS record number and the associated MAS sequence number from the cross-reference field. Some of the cross-references found in the MRDS data had already been found in the MAS cross-reference information. Those cross-references which had been previously identified within the MAS database were not used and only the unique cross-references found in MRDS were added to the ASCII cross-reference file (Table 4). This file was then imported into ARC/INFO format and linked to both the MAS and MRDS ARC/INFO files. This link provided the means to create

subset files of each database that contained only those records with a cross-reference to the other database.

It was desired to evaluate the cross-reference between the two databases by determining what information for each cross-referenced pair was identical. The following fields were compared for each cross-referenced pair of mines: state, county, latitude, longitude, commodity name, commodity modifier, watershed and site name. Four other fields were considered for comparison and rejected. Elevation information was limited in both databases and it was thought that comparing elevation would not be of significant value. Development status was the type of information that would depend heavily on the time of data collection and therefore not used for comparison. Workings type in MRDS was very general and thought to not be of significant value. Finally, alternate name was not compared because it was desired to determine the outcome of the site name comparisons before considering alternate name. Additional data fields were added to the MRDS ARC/INFO cross-referenced subset file to contain the results of the comparisons for each of the above fields. The operations performed (within ARC/INFO except where noted) in order to compare the information for each MRDS mine to the information for the corresponding MAS mine are described below:

1) State. The MRDS state code (a 2 character alpha code) was compared to the MAS state code (the first three digits of the Federal Information Processing System (FIPS) code, with the first digit always zero). In order to allow for this comparison, an ARC/INFO lookup table was created that contained each 2 character MRDS alpha state code, the corresponding three digit MAS FIPS state code, and the state name. A 'state code match' field on each MRDS record was updated with a '1' for those records with state codes that matched the MAS state code and a '0' for those records that did not match the MAS state code (Table 5). For those records that did not have a matching state code, a test was conducted to determine if either the MAS FIPS code or the MRDS alpha code was transposed. Also for these records, the possibility of determining if the state codes represented adjacent states was considered but rejected due to the effort that would be involved for the small amount of information gained.

	Exact Match	Non-Match	
State	21319	25*	
County	21071#	273##	
Commodity name only:			
1st MRDS to any MAS	19404	1940	
Any MRDS to any MAS	19899	1445	
Commodity name and modifier:			
1st MRDS to any MAS	14270	7074	
Any MRDS to any MAS	16186	5158	
Watershed	20279	1065	
Site Name	15872	5472	
	Yes	**No**	
Latitude within 1 minute	19502	1842	
Longitude within 1 minute	19910	1434	
	Equals 0	**> 0 and <= 1**	**> 1**
Distance in Kilometers	7874	9307	4163

*0 records with transposed state codes.
#37 records where one of two MRDS county names matched MAS county name.
##42 records with 'SAINT' in MRDS county name but 'ST' in MAS county name, 0 misspelled county names.

Table 5. Summary of cross-referenced subset comparison results.

2) County. The MRDS county name was compared to the MAS county code (the last three digits of the FIPS code). In order to allow for this comparison, an ARC/INFO lookup table was created that contained each county name and the corresponding three digit MAS FIPS county code. A 'county match' field on each MRDS record was updated with a '1' for those records with county names that matched the MAS county code and a '0' for those records that did not match the MAS county code. For those records that did not have a matching county, a list of the MRDS county names and the corresponding MAS county codes (and county names from the lookup table) was evaluated to determine if any of the county names were simply misspelled (Table 5). Evaluating the possibility of transposed MAS FIPS county codes for these records was considered but rejected due to the large number of possible combinations that would need to be tested for the 3 digit code. Also for these records, the possibility of determining if the

counties represented adjacent counties was considered but rejected due to the effort that would be involved for the small amount of information gained.

3) Latitude. The MRDS latitude was compared to the MAS latitude. It was first necessary to copy the coordinate information for each MRDS and MAS record from the MRDS and MAS ARC/INFO spatial data files to the appropriate ARC/INFO attribute data files using the 'addxy' command. Then the difference between the latitude of each MRDS mine record and the latitude of the corresponding MAS mine record was calculated. Next, the 'latitude within one minute' field on each MRDS record was updated with a '1' for those records with a latitude that was different from the MAS latitude by one minute or less. Similarly, this field was updated with a '0' for those records with a latitude that was different from the MAS latitude by more than one minute (Table 5). Finally, a 'latitude match' field on each MRDS record was updated with a '1' for those records with a latitude that exactly matched the MAS latitude and a '0' for those records with a latitude that did not match the MAS latitude.

4) Longitude. The MRDS longitude was compared to the MAS longitude. The same steps as described above were performed for longitude with the fields 'longitude within one minute' and 'longitude match' updated appropriately (Table 5).

5) Commodity. The MRDS commodity information was compared to the MAS commodity information. First the commodity names only and then the commodity names and modifiers were compared. An ARC/INFO lookup table had already been created for the MAS database which contained the 4 digit MAS commodity code and the corresponding commodity name and modifier. It was necessary to add the appropriate 1 to 4 character MRDS commodity code to this lookup table in order to compare commodity names and modifiers. Each database could contain 1 or more commodity codes per record. The MRDS commodity codes were recorded on each record in order of the importance of the commodity produced at the mine. The MAS commodity codes were not recorded in any such order. Therefore the comparisons began by comparing the first MRDS commodity to any MAS commodity. A 'commodity name match' field on each MRDS record was updated with a '1' if the name for the first MRDS commodity matched the name for any MAS commodity, otherwise the field was updated with a '0'. Similarly, a 'commodity name & modifier match' field on each MRDS record was updated with a '1' if the name and modifier for the first MRDS commodity matched the name and modifier for any MAS commodity, otherwise the field was updated with a '0' (Table 5).

It was then decided that it would be helpful to know if *any* MRDS commodity matched any MAS commodity. Therefore two additional match fields were added to MRDS to hold the results of the 'any commodity name match' and 'any commodity name & modifier match' comparisons to MAS. These fields were updated in a manner similar to that described above except the name (and modifier) for *any* MRDS commodity was compared to the name (and modifier) for any MAS commodity (Table 5).

6) Watershed. It was desired to determine if each MRDS mine occurred in the same 8-digit hydrologic unit (i.e., watershed) as the corresponding MAS mine. In order to determine this the

MRDS and MAS ARC/INFO cross-referenced subset files were projected into the Albers projection in order for the files to be in the same projection as the ARC/INFO file that contained the watershed boundaries for the conterminous U.S. An ARC/INFO 'identity' was performed between each mines file and the watershed boundaries file in order to determine the watershed in which each mine occurred. A number that designated the watershed was added to each record in each of the mines files as part of this process. It was then possible to determine if each cross-referenced pair of mines occurred in the same or different watersheds. A 'watershed match' field on each MRDS record was updated with a '1' for those mines that occurred in the same watershed as the MAS mine and a '0' for those mines that did not occur in the same watershed as the MAS mine. For those mines where either the MRDS or MAS record occurred in a watershed outside the boundaries of the conterminous U.S. the 'watershed match' field was updated to a '0' (Table 5).

7) Site Name. The MRDS site name was compared to the MAS site name first checking for an exact match. A 'name match' field on each MRDS record was updated with a '1' for those records with site names that exactly matched the MAS site name and a '0' for those records that did not match the MAS site name (Table 5). Since site name was considered to be a very important field and since even minor differences in spelling could cause an MRDS site name to not exactly match an MAS site name additional comparisons were completed for all records with a 'name match' field initially set to '0'. Initially an attempt was made to compare by hand those records that did not match exactly on site name. It was then decided that besides being time consuming and tedious this method was subject to bias when determining if two names that seem similar actually were the same or not. Therefore, for all pairs that did not have an exact match on site name, a C program was written to systematically compare the MRDS and MAS site names. This was a substantial effort that evolved as knowledge was gained about the site name data. Several preliminary steps were developed and programmed that were not used in the final version of the comparison program (Appendix A). The MRDS identification number, MRDS site name string and corresponding MAS site name string were exported into an ASCII file in order for the final version of the comparison program to read each pair of names and perform the various functions to compare the two name strings. The steps are described in the order in which they were performed:

a) All punctuation marks were deleted from each name string and each string was reformatted so each word was separated by only one blank space.

b) All extraneous words were deleted from each name string. Extraneous words were words such as 'the', 'and', 'of', as well as generic names such as 'mine', 'pit' and 'mill'. Words that denoted that a site name was not known such as 'unknown' or 'unnamed' were also considered extraneous words and therefore deleted. In many cases deletion of these types of words would cause a name string to become a null string (see item d below). A complete list of the extraneous words is provided in Appendix B.

c) All known abbreviations within each string were replaced with their appropriate non-abbreviated words, Roman numerals (I through X) were replaced with their appropriate Arabic

numerals, and the words 'one' through 'ten' were replaced with their appropriate Arabic numerals (Appendix C).

d) If either name string after the first three steps had become a null string, the site names were considered to be non-matches, a '0' was output to an ASCII file to denote that the strings did not match and no more processing was performed on these strings. If both name strings were not null and they matched exactly at this point, a '1' was output to an ASCII file to denote this, and no more processing was performed on these strings. Otherwise, processing continued.

e) All diphthongs in all words in each string were replaced with a unique non-letter ASCII character (Appendix D), and all occurrences of double letters were replaced with the appropriate single letter. After this project was completed it was found that the diphthongs 'HY' and 'PR' were both replaced with the '@' character (Appendix D). This would potentially be a problem if one of these diphthongs was found in one of the site name strings and the other diphthong was found in the other site name string. A search was conducted for this condition among the group of MRDS and MAS site name strings that did not have an exact match and this condition was not found.

f) All vowels were removed from all words in each string.

g) All spaces were removed from each string and the strings were compared. If the strings matched exactly, a '1' was output to an ASCII file to denote this and no more processing was performed on these strings, otherwise processing continued with the strings as they were before the spaces were removed.

h) Duplicate words within each name string were deleted.

i) Each word in the MRDS name string was compared to each word in the MAS name string to determine if the name strings were composed of the same words, but simply in different orders. If this was the case, the strings were considered exact matches, a '1' was output to an ASCII file to denote this and no more processing was performed on these strings. If the name strings still did not match the processing continued.

j) Finally, a word by word, letter by letter, between string comparison was conducted first comparing the name string with the larger number of words to the name string with the smaller number of words and then vice versa. As an example, consider an MRDS name string with 3 words and an MAS name string with 2 words:

STRING 1 (MRDS)			STRING 2 (MAS)	
WORD A	WORD B	WORD C	WORD D	WORD E

First, each word in STRING 1 was compared to each word in STRING 2. Word A was compared letter by letter to both words D and E and the results of the better of the two matches (A to D or A to E) were retained in order to calculate the overall percent match of STRING 1 to STRING 2. Words B and C were compared to both words D and E in a like manner and then the overall match percentage of STRING 1 to STRING 2 was calculated.

The method for comparing words letter by letter will be described first and then the method for calculating the overall string match percentage will be outlined. The following example illustrates the method used for comparing words:

WORD A	WORD D
ALHAMBRA	ALLUMBRA

Each letter in WORD A was compared to all letters in WORD D (always beginning with the first letter in WORD D) until a matching letter in WORD D was found or the end of the word was reached. If a matching letter in WORD D was found, that letter was replaced with a blank so no more letters in WORD D could match it. The total number of letters in WORD A that have a matching letter in WORD D was calculated, and the percent match of WORD A to WORD D was found by dividing this total number of letters matched by the total number of letters in WORD A. This method is illustrated step by step for the above two words:

STEP 1:	ALHAMBRA ^	ALLUMBRA ^	# Letters Match = 1
STEP 2:	ALHAMBRA ^	LLUMBRA ^	# Letters Match = 2
STEP 3:	ALHAMBRA ^	LUMBRA	No Match for 'H' # Letters Match = 2
STEP 4:	ALHAMBRA ^	LUMBRA ^	# Letters Match = 3
STEP 5:	ALHAMBRA ^	LUMBR ^	# Letters Match = 4
STEP 6:	ALHAMBRA ^	LU BR ^	# Letters Match = 5
STEP 7:	ALHAMBRA ^	LU R ^	# Letters Match = 6
STEP 8:	ALHAMBRA ^	LU	No Match for 'A' # Letters Match = 6

$$\text{Word Percent Match of WORD A to WORD D} = \frac{\text{\# Letters Match}}{\text{Total \# letters in WORD A}} = \frac{6}{8} = .75$$

Once each word in STRING 1 was compared to each word in STRING 2 and the best word percent match for each word in STRING 1 was determined the overall string percent match was then calculated. The best word percent match for each of the words in STRING 1 was first divided by the number of words in STRING 1 and the results were then summed to produce an overall percent match of STRING 1 to STRING 2. This can be illustrated as follows:

STRING 1 Overall % Match = [(Maximum of WORD A as % of WORD (i)) / 3] +
[(Maximum of WORD B as % of WORD (i)) / 3] +
[(Maximum of WORD C as % of WORD (i)) / 3]

(Where i = D, E)

Finally, the name string with the smaller number of words (STRING 2) was compared to the name string with the larger number of words (STRING 1) in the same manner and the overall string percent match was calculated. This can be illustrated for the example as follows:

STRING 2 Overall % Match = [(Maximum of WORD D as % of WORD (j)) / 2] +
[(Maximum of WORD E as % of WORD (j)) / 2]

(Where j = A, B, C)

The output of the comparison program included a MRDS identification number, a '1' or '0' to denote if the name strings matched exactly or not, and for strings which did not match exactly, the string overall percent match for the longer string compared to the shorter string and vice versa. This information was imported into ARC/INFO and linked to the MRDS cross-referenced subset file (Table 5).

III. DEVELOPING PROCEDURES FOR IDENTIFYING ADDITIONAL DUPLICATES

Prior to completion of the steps outlined above for comparing site name and evaluating watershed between cross-referenced records (section II-6,7) it was decided that it would be necessary to develop a procedure that could be used to evaluate those records in the two databases that did not have an explicit cross-reference. A procedure was needed for determining those characteristics shared by a pair of mines (one from each database) which identified a pair as duplicates. Once the characteristics and their relative importance were determined then the two databases could be searched for those pairs of mines that met the criteria (i.e., duplicate

mines). It was decided that a logistic regression procedure would be used to identify the important characteristics of a duplicate pair of mines.

The logistic regression would need as input a number of pairs of mines that were known to be duplicates and a number of pairs that were not duplicates. The "known duplicates" would be obtained from the subset of MAS and MRDS mines that had an explicit cross-reference linking records from one database to their appropriate corresponding record in the other database. The "known not duplicates" would be obtained by selecting pairs of mines (without cross-reference information) from within each database after each database was first checked for duplicate records within. Five thousand pairs of "known duplicates" and "known not duplicates" would be randomly selected from the appropriate databases to create a set of 10,000 observations for input into the logistic regression.

Many of the operations described in section II were used to generate the set of independent variables needed for the logistic regression. The results of the state, county, commodity name, commodity name and modifier, and watershed comparisons would be entered into the logistic regression as either a '0' to denote a non-match or a '1' to denote a match for each item compared. For comparisons of the site name the result of the name match comparison would be entered as either a '0' or a '1'. Additionally for those site names that did not match exactly the overall percent matches of the name string with the larger number of words compared to the name string with the smaller number of words and vice versa would be entered as a decimal percent between 0 and 1. A distance measure in kilometers of the ground distance that separates each pair of mines would be entered rather than the results of the latitude and longitude comparisons described in section II-3,4. (The steps for performing the distance measurements will be described in section IV.) Finally, identification information for each pair of mines and the value of the dependent variable ('1' for "known duplicates" and '0' for "known not duplicates") would be entered into the logistic regression.

IV. DISTANCE CALCULATIONS

The distance measurement calculations for the cross-referenced subsets of mines were performed after first projecting the records into the UTM projection in order to measure distance in meters rather than degrees. The UTM system divides the earth into 60 zones, 10 of which cover the conterminous U.S. The MRDS cross-referenced subset of mines was subdivided into groups based on the longitude values of each mines' location and then each group was projected into the appropriate UTM zone. The same procedure was performed on the MAS cross-referenced subset of mines. A problem developed when some MRDS mines in a particular UTM zone were found to be cross-referenced to MAS mines in a different UTM zone. The distance between mine records occurring in different zones could not be calculated.

Two attempts were made to resolve the problem of cross-referenced mine records occurring in different zones. First, the MAS cross-referenced subset of mines was subdivided into groups based on the longitude values of each mines' location with a one-quarter degree buffer zone to

the east and west of each UTM zones' boundary. Each group was then projected into the appropriate UTM zone. However the same problem occurred but with fewer cross-referenced mines. Finally, the entire MAS cross-referenced subset of mines was projected into *each* UTM zone. This allowed the MRDS mines in each zone to be linked to the appropriate MAS mines all of which were also projected into the same zone. It was then possible to actually calculate the distance between each MRDS mine and its cross-referenced MAS mine.

It was first necessary to copy the coordinate information (in units of meters for the UTM projection) for each MRDS and each MAS record in each zone from the ARC/INFO spatial data file to the ARC/INFO attribute data file. This was accomplished using the ARC/INFO 'addxy' command. Next, 3 new fields were added to each MRDS record in each zone. The fields were "X difference", "Y difference" (difference in meters between each MRDS mine and its cross-referenced MAS mine in the north-south and east-west directions, respectively) and "distance". The "X difference" was calculated by subtracting the MAS mine easting value from the appropriate MRDS mine easting value. The "Y difference" was calculated by subtracting the MAS mine northing value from the appropriate MRDS mine northing value. If any of the difference values were negative they were multiplied by -1 to make them positive. The "distance" was then calculated using the Pythagorean theorem: the square root of the sum of the squares of the "X difference" and "Y difference". Finally a fourth new field was added to each MRDS record in each zone to hold the "distance in kilometers" value and set equal to the "distance" divided by 1000. In this manner all mines in the MRDS cross-referenced subset were updated with the appropriate values of the distance to the corresponding mine in the MAS cross-referenced subset (Table 5).

V. REMOVING DUPLICATES FROM CROSS-REFERENCED SUBSET

Before building the data set of 5000 "known duplicates" to be used in the logistic regression it was necessary to ensure that each mine in the subset of mines from each database that contained a cross-reference was linked to one and only one mine in the other database. A frequency analysis using SAS was performed on the MAS and MRDS cross-reference numbers. In those cases where more than one MRDS mine was linked to the same MAS mine or vice versa the mine record which contained the greatest number of commodity codes was retained and all other mine records with the same cross-reference number were deleted from this cross-referenced subset. The cross-reference number for these mines was also deleted from the original MAS and MRDS ARC/INFO files that contain all the mine records.

VI. BUILDING NON-CROSS-REFERENCED SUBSET AND REMOVING DUPLICATES

It was previously described how subsets of the two databases were created for those records with cross-reference information. This cross-reference information also enabled creation of subset files of each database that contained only those records *without* any cross-reference information between the two databases (Table 6). It was from these files that pairs of "known not duplicates" would be chosen for the logistic regression. Once files were created for both the MAS and

MRDS databases of all records without cross-reference information, the files were further divided so all records occurring in Hawaii, Alaska, and the conterminous U.S. were in separate files (Table 6). The Alaska and U.S. files were then projected into the Albers projection. This enabled an ARC/INFO 'identity' command to be performed between each U.S. and Alaska file and the appropriate watershed boundaries file in order to determine in which watershed each mine occurred. The Hawaii files were projected into UTM zone 4. For Hawaii, the 8-digit hydrologic units correspond to the individual islands. Finally, the coordinate information (in meters for both the Albers and UTM projections) for each MRDS and each MAS record was copied from the ARC/INFO spatial data file to the ARC/INFO attribute data file using the ARC/INFO 'addxy' command.

	Original Count	Duplicates Found & Deleted	Final Count	Pairs Selected for Logistic Regression
MRDS records in non-cross-reference subset file	48158	314	47844	1025
Portion in conterminous U.S.	44035	310	43725	937
Portion in Alaska	4120	4	4116	88
Portion in Hawaii	3	0	3	0
Mas records in non-cross-reference subset file	185898	444	185454	3975
Portion in conterminous U.S.	178581	435	178146	3818
Portion in Alaska	7262	8	7254	156
Portion in Hawaii	55	1	54	1

Table 6. Summary of records without cross-reference information.

A C program was written to search for duplicate records *within* each of the 3 MRDS and 3 MAS subset files. Appropriate fields from each record in the 6 files were exported to an ASCII file. The fields exported were: MRDS or MAS identification code, state code, county code (county name for the MRDS files), X coordinate, Y coordinate, watershed, commodity codes, and site name. The program was run 6 times - once for each of the 6 files. The first step in the program was to preprocess the site names for all records. This preprocessing included many of the same steps described in the site name comparisons in section II-7. All punctuation marks and extraneous words were deleted from each name string. All abbreviations were replaced with the appropriate non-abbreviated words. Duplicate words within each string were deleted. All

diphthongs were replaced with unique non-letter ASCII characters. All vowels and spaces were removed. The resulting name string was then written to a new ASCII file for later use.

The second step in the program was to compare the information for each record in the ASCII file to the information for every other record within the same file in order to find the record with the largest number of fields that match. The process was to read a record and save the values for each of the 8 fields noted above. Then every other record in the file was read and a direct comparison was made between the 8 fields of the originally read record and the 8 fields of every other record. Each comparison was a simple comparison for equality and the results of each comparison were saved. For example, if the state code field on the originally read record was equal to (exactly the same as) the state code field on another record, the state codes were considered matches, otherwise they were not. The same held true for the remaining 7 fields except site name. If a site name was null it was not compared to the site names from other records and therefore would not be considered a match to any record's site name. Otherwise the site name was compared for exact equality (i.e., no string match percentages were calculated in this program). The number of fields that matched for each record comparison was evaluated and the information for the current largest number of fields matched was maintained. This continued until all records were read and compared to the originally read record. At this point the results of each field comparison ('0' or '1' for nonmatch or match) for the record with the largest number of matching fields as well as the identification numbers for both records were written to an ASCII file, and the process was repeated for the next record..

The new ASCII file created in the second step contained the results of the 8 field comparisons for all records in the file. There results were then evaluated in a spreadsheet to determine which records in each of the 6 files had a record that could be considered a duplicate. Some rules were developed when evaluating the results:

1) The site names for a given pair of records must match exactly (the field comparison result for site name must equal '1') in order for a pair to be considered a duplicate.

2) If either the X coordinates or Y coordinates of a given pair of records match exactly then the X and Y coordinates and watershed for this pair of records are considered to match exactly.

3) If the X and Y coordinates and watershed of a given pair of records match exactly then the state and county for this pair of records are considered to match exactly.

Once these rules were established it was possible to build a table of all combinations of field match values that would denote that a pair of records should be considered an exact match and therefore one of the records in the pair was a duplicate and could be deleted from the data set. For example, if the field match values for a given pair of records for state, county, X coordinate, Y coordinate, watershed, commodity, and site name were all equal to 1 then this pair would be an exact match. This case is one of 41 cases that define an exact match as shown in Appendix E. The values in the table in Appendix E were used as a filter in the spreadsheet to process the ASCII results file generated by the C program for each of the 6 files in order to find those pairs

of records that represented duplicates. The highest identification number of each pair that contained a duplicate was written to a new file and used to build an ARC Macro Language (AML) program to delete this duplicate record from the appropriate ARC/INFO file (Table 6).

VII. BUILDING LOGISTIC REGRESSION DATA SET

As mentioned in section III, 5,000 pairs of "known duplicates" and "known not duplicates" would be randomly selected from the appropriate databases to create a set of 10,000 observations for input into the logistic regression. The 5,000 pairs of "known duplicates" were selected from the MRDS and MAS cross-referenced subset files. In order to randomly select 5,000 pairs, a C program was written to create an ASCII file that contained the same number of records as the MRDS and MAS cross-referenced subset files. Each record created contained a record number (beginning with 1 and incremented by 1 for each record created) and a random number. This ASCII file was then imported into ARC/INFO and linked to the MRDS cross-referenced subset file. Each record in the MRDS cross-referenced subset file then contained a field with a random number. All MRDS records which contained a random number below a certain cut-off were selected as were the related MAS records in order to choose the 5,000 pairs. Finally the fields to be used in the logistic regression (described in section III) for each of the 5,000 pairs were exported into an ASCII file.

The 5,000 pairs of "known not duplicates" would be randomly selected from the 6 non-cross-referenced subset files, proportionally (Table 6). Unlike the 5,000 "known duplicate" pairs, the "known not duplicate" pairs would be chosen from *within* each file. In other words, once a record was randomly selected from a file the second record in the pair would be chosen from the same file by searching for a record within 1 kilometer of the location of the originally selected record. The 1 kilometer distance was chosen after evaluating a frequency distribution of the distances separating the MRDS and MAS cross-referenced subset records. An AML was written to control selecting the pairs of records from within each file, performing the appropriate comparisons and exporting the results of the comparisons into an ASCII file in order to be used in the logistic regression.

The AML first accessed the program described above to create an ASCII file with each record containing a record number and a random number. This file was then imported into ARC/INFO and sorted by the random number. A new ASCII file was created by exporting only the record numbers (now sorted in a random order). The record numbers in this file were used to randomly select a record from the appropriate non-cross-referenced subset file. At this point in the AML a loop began that continued until the correct number of pairs of records were selected from the appropriate file. The ARC/INFO 'reselect' command was used to create a new ARC/INFO file that contained only the 1 currently selected record. A second new ARC/INFO file was then created using the 'buffer' command to build a 1 kilometer buffer area around the currently selected record. This buffer area file was used to 'clip' all records that fell within the buffer area from the original non-cross-referenced subset file and place them in a third new file. If no records fell within the buffer area, processing would transfer back to the loop beginning to

randomly select the next record. Otherwise, each record in the buffer area was paired with the originally selected record and the necessary comparisons were performed for each of these pairs.

Checks for exact matches between the pairs were performed for state, county, watershed, any commodity name, and any commodity name and modifier. The results of these comparisons were written to an ASCII file as either a '1' (match) or '0' (non-match). The distance in kilometers that separated each pair was calculated by first determining the "X difference" and "Y difference" and then using the Pythagorean theorem as described in section IV. A check for an exact match between the site names was performed and for those pairs that did not have site names that matched exactly a C program was accessed in order to complete the same name comparisons as those described in section II-7. Once all the comparisons were completed, the results along with identification information for each pair and the value of the dependent variable ('0' for these "known not duplicates") were written to an ASCII file to be used in the logistic regression (as described in section III). The AML processing then returned to the start of the loop in order to select the next record. This continued until 5,000 pairs of "known not duplicates" were chosen and processed.

VIII. PERFORMING LOGISTIC REGRESSION

Once the regression dataset had been created a simple linear logistic regression analysis was performed. The goal of this analysis was to develop a model that could predict the probability that a given pair of mine records (one from MAS and one from MRDS) actually described the same mine. The form of this model was:

$$\log(p / 1 - p) = \alpha + \beta x$$

Where p is the probability that a pair of mine records describe the same mine, α is an intercept parameter, β is a vector of slope parameters and x is a matrix of explanatory variables.

The explanatory variables used in the model were selected using stepwise regression procedures with both the tolerance values for entering new variables into the model and for leaving existing variables in the model set at 0.05. Nine variables were originally considered for possible inclusion in the stepwise regression analysis (table 7), but some of these explanatory variables were highly correlated with one another and thus would have created an autocorrelation problem if they appeared together in the regression model. Consequently, a simple correlation analysis was used to identify two groups of explanatory variables that were highly correlated with one another (person correlation coefficients of at least 0.50 in absolute value). The single explanatory variable from each group that was most highly correlated with the probability of a pair of mine records describing a single mine (the dependent variable) was identified and retained for use in the stepwise regression analysis, and the remaining explanatory variables in each group were dropped (Table 8).

State	County	Commodity Name	Commodity Name and Modifier
Watershed	Site Name	Short Site Name	Long Site Name
	Distance		

Table 7. Explanatory variables considered for use in logistic regression analysis

	Explanatory Variable	Correlation with Dependent Variable	Included in the Stepwise Regression Analysis
Group 1:	Site Name	0.6047	
	Short Site Name	0.7281	✓
	Long Site Name	0.7234	
Group 2:	Commodity Name	0.1938	✓
	Commodity Name and Modifier	0.0293	

Table 8: Groups of Highly Correlated Explanatory Variables

This resulted in six potential explanatory variables for inclusion in the stepwise regression analysis. Five of these variables were ultimately entered into the model. Full details of the regression analysis are included in appendix F, but a summary of the final results are shown in table 9.

Variable	Degrees of freedom	Parameter Estimate	Standard Error	Wald χ^2	*p*-Value
Intercept	1	-3.4248	0.5050	45.9982	0.0001
County	1	1.2221	0.3544	11.8886	0.0006
Commodity	1	0.8912	0.1215	53.7689	0.0001
Watershed	1	-0.7600	0.3408	4.9727	0.0257
Short Name	1	5.3007	0.1093	2,351.8317	0.0002
Distance	1	-3.8081	0.1219	976.5818	0.0001

Concordant	94.5%	Somers' D	0.893
Discordant	5.2%	Gamma	0.896
Tied	0.3%	Tau-a	0.441
		c	0.947

Table 9: Results of Stepwise Regression Analysis

It should be noted that when the stepwise regression analysis was first performed on the full dataset consisting of 5,000 matched and 5,000 unmatched pairs of mines the results produced were very counter intuitive: The probability of a pair of mine records describing the same mine was predicted to *increase* as the distance between the two mines increased. An investigation of this phenomenon revealed that it was being caused by a relatively small percentage (approximately 20%) of the 5,000 matched mine records that had extremely large distance measures. Consequently, all matched mine pairs with distance measures greater than 1 kilometer were eliminated from the dataset and a new stepwise regression analysis was conducted. This procedure resulted in a dataset consisting of 5,000 unmatched pairs of mines and 3,978 matched pairs of mines. The results of this second stepwise regression analysis were intuitively reasonable. It is these results from the second regression analysis that are presented here.

This would indicate to me that the latitude/longitude measures are not very reliable.

IX. APPLYING RESULTS OF REGRESSION

Once the proper coefficients were determined it was then necessary to apply these coefficients to every reasonable pair of MRDS and MAS records in order to find the MAS record which had the highest probability of a match to each MRDS record. In order to reduce processing time the MRDS and MAS files for the conterminous U.S. were subset into 6 files each with

approximately equal numbers of records in the MAS subset files. This breakdown occurred by state such that all mine records for a given state were within one of the subset files. The MRDS and MAS files for Alaska and Hawaii were processed separately.

An AML was created to control processing of all records within an MRDS subset file to compare to the records in the proper MAS subset file. The first step was to select a record from the MRDS file and find all MAS records that were within 1.05 km of the MRDS location. This was done to drastically reduce the number of MAS records to which each MRDS record would be compared. The 1.05 km distance was chosen since the regression coefficient for distance showed that an MAS record farther than 1.05 km from an MRDS record would have a probability of matching the MRDS record of 0 percent. Next, the MRDS record was compared to each MAS record to determine if the records' county, watershed, and any name commodity matched. The distance in kilometers between the records was also calculated. Finally, the site names of the records were compared to determine to what degree they matched. For those site names that did not match exactly a C program was accessed in order to complete the same name comparisons as those described in section II-7. The results of these comparisons were then used in another C program that calculated the probability (between 0 and 1) of the two records' being a match using the appropriate regression coefficients. The MAS Sequence Number of the MAS record which had the highest probability value of matching the MRDS record and the probability value were saved as new items on the MRDS record. (Note: it was possible that no MAS records were located within 1.05 km of the MRDS record. In this case the MAS Sequence Number and probability value were both recorded as 0 on the MRDS record.) The next MRDS record was then selected and the above steps were repeated until all 47844 MRDS records were processed for all conterminous U.S. subset files and for the Alaska and Hawaii files.

RESULTS

The probability values of 47844 MRDS records were evaluated using a frequency procedure in order to determine the number of records with probabilities greater than zero that occurred in each of 20 classes. Probabilities greater than 0 and less than or equal to 1 were divided into 20 equal classes with a width of 0.05 (figure 1). Figure 1 does not include the 14063 MRDS records that had a 0 probability of matching a record in the MAS file. A large number of MRDS records (11332, or 24%) had probabilities of matching records in the MAS file of greater than 0.90.

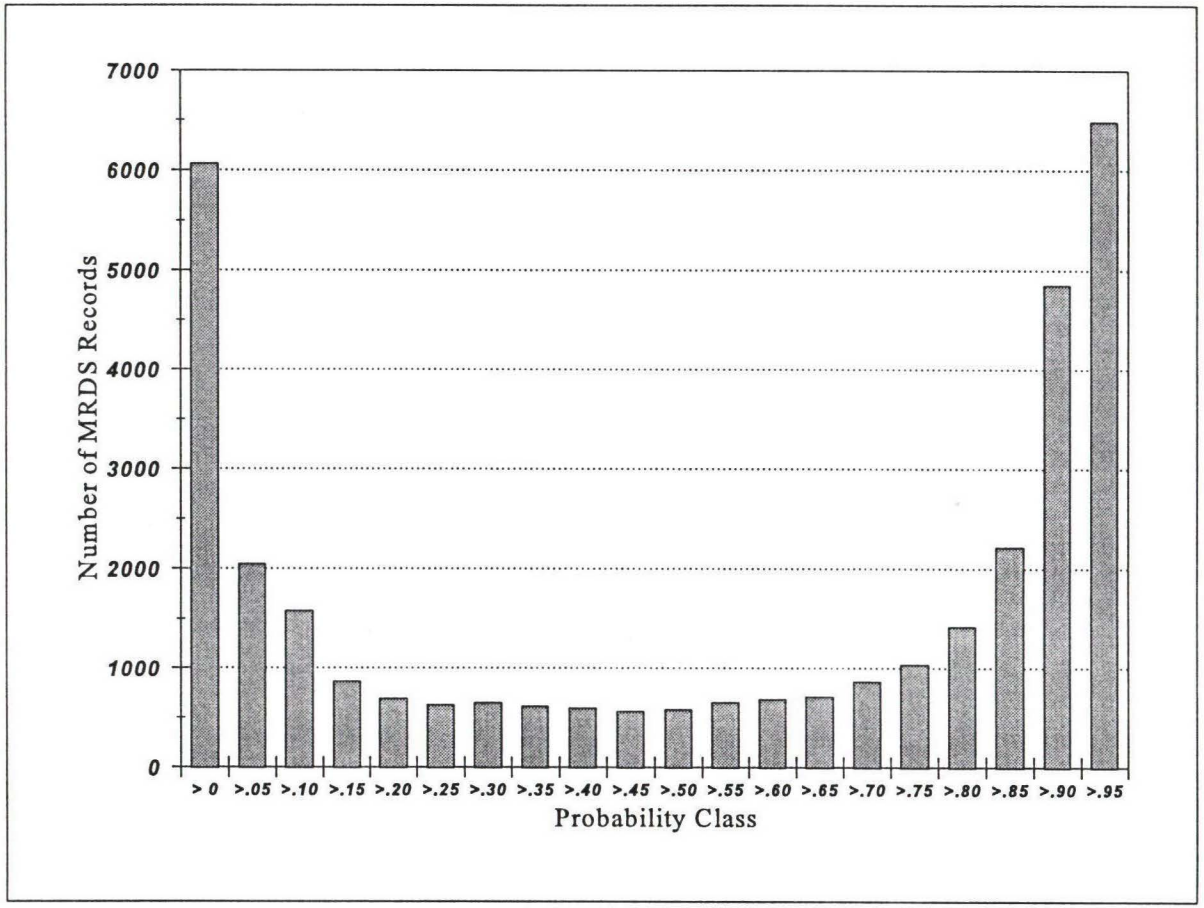

Figure 1. Summary of MRDS probabilities greater than zero.

Appendix A. Preliminary site name comparison steps.

Several steps were developed and programmed for site name comparisons that ultimately were not used in the final version of the comparison program. These steps were helpful in evaluating the data and learning about the information contained in the site name fields. The steps were developed to find site name strings that met certain conditions and are summarized here:

1) Strings with only 1 "word" and the "word" is a digit or group of digits.

2) Strings with only 1 word and the word is a commodity or commodity-like word (e.g. "gravel").

3) Pairs of strings with all words matching but one or more numbers not matching.

4) Pairs of strings with all numbers matching but one or more words not matching.

5) Pairs of strings with all words in the MRDS string matching words in the MAS string but at least 1 word in the MAS string not matching a word in the MRDS string.

6) Pairs of strings with all words in the MAS string matching words in the MRDS string but at least 1 word in the MRDS string not matching a word in the MAS string.

7) Pairs of strings with at least one word from each string not matching any word in the opposite string and more MRDS words matching than not matching words in the MAS string.

8) Pairs of strings with at least one word from each string not matching any word in the opposite string and more (or the same number of) MRDS words not matching than matching words in the MAS string.

Appendix B. Extraneous words.

ADIT	EXTENTION	OCCURRENCES	SHAFTS
ADITS	FIND	OCCURRANCE	STRIKE
&	FINDS	OF	STRIKES
AND	GROUP	ON	THE
AREA	GROUPS	OPERATION	THROUGH
BODY	INC	OPERATIONS	THRU
CLAIM	INCORPORATED	ORE	TUNNEL
CLAIMS	LOCATION	PATENT	TRACT
CONS	MILL	PATENTS	UNDETERMINED
CONSOLIDATED	MINE	PIT	UNIDENTIFIED
CO	MINES	PITS	UNK
COMPANY	MINING	PLANT	UNKNOWN
COMPANIES	MAG	PLANTS	UNNAMED
CORP	NAME	PORPSECT	UNNAAMED
CORPS	NAMELESS	POSSIBLE	VEIN
DEP	#	POSSIBLY	WELL
DEPOSIT	NO	PROJECT	WELLS
DEPOSITS	NOS	PROJECTS	WORK
DEPOSIIT	NUM	PROP	WORKS
DEV	NUMS	PROPERTY	WORKING
DEVELOPMENT	NUMBER	PROPERTIES	WORKINGS
DISTRICT	OCC	PROSPECT	
EXT	OCCURENCE	PROSPECTS	
EXTENSION	OCCURRENCE	SHAFT	

Appendix C. Abbreviations and replacements.

Abbreviation	Replacement Word	Word/Roman Numeral	Replacement Numeral
BLDG	BUILDING	ONE	1
BROS	BROTHERS	TWO	2
BRS	BROTHERS	THREE	3
CALIF	CALIFORNIA	FOUR	4
CHEIF	CHIEF	FIVE	5
CONST	CONSTRUCTION	SIX	6
CRK	CREEK	SEVEN	7
CR	CREEK	EIGHT	8
DEPT	DEPARTMENT	NINE	9
DR	DOCTOR	TEN	10
FK	FORK	I	1
FT	FORT	II	2
GREY	GRAY	III	3
HWY	HIGHWAY	IV	4
MT	MOUNTAIN	V	5
MTN	MOUNTAIN	VI	6
MOUNT	MOUNTAIN	VII	7
MOUNTAINS	MOUNTAIN	VIII	8
N	NORTH	IX	9
QTZ	QUARTZ	X	10
ST	SAINT		

Appendix D. Diphthongs and replacement characters.

Diphthong	Replacement Character
SH	[
CH	]
PH	{
GH	}
WH	^
QU	<
TH	>
HY	@
ST	~
TR	%
PR	@
CK	!

Appendix E. Field match values that define exact matches.

State	County	X Coord	Y Coord	Watershed	Commodity	Site Name
1	1	1	1	1	1	1
1	1	1	1	1	0	1
1	1	1	1	0	1	1
1	1	1	0	1	1	1
1	1	0	1	1	1	1
1	0	1	1	1	1	1
0	1	1	1	1	1	1
1	1	1	0	1	0	1
1	1	0	1	1	0	1
1	0	1	1	1	0	1
0	1	1	1	1	0	1
1	1	1	0	0	1	1
1	1	0	1	0	1	1
1	0	1	0	1	1	1
0	1	1	0	1	1	1
1	0	0	1	1	1	1
0	1	0	1	1	1	1
0	0	1	1	1	1	1
1	1	1	0	0	0	1
1	1	0	1	0	0	1
1	0	1	0	1	0	1
0	1	1	0	1	0	1
1	0	0	1	1	0	1
0	1	0	1	1	0	1
0	0	1	1	1	0	1
1	0	1	0	0	1	1
0	1	1	0	0	1	1
1	0	0	1	0	1	1
0	1	0	1	0	1	1
0	0	1	0	1	1	1
0	0	0	1	1	1	1
0	0	0	1	0	1	1
0	0	1	0	0	1	1
0	0	0	1	1	0	1
0	0	1	0	1	0	1
0	1	0	1	0	0	1
1	0	0	1	0	0	1
0	1	1	0	0	0	1
1	0	1	0	0	0	1
0	0	0	1	0	0	1
0	0	1	0	0	0	1

Appendix F. Detailed Results of Linear Logistic Regression Analysis

Variables Used:

SMC	State match code (1 if states match; 0 otherwise)
CMC	County match code (1 if counties match; 0 otherwise)
CNMC	Commodity name match code (1 if commodities match; 0 otherwise)
WMC	Watershed match code (1 if watersheds match, 0 otherwise)
SNP	Short name percentage (percentage of match between names based on shorter name)
Dist	Distance (in meters) between recorded mine locations

Stepwise Selection Procedure

Step 0. Intercept entered:

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald χ^2	Pr > χ^2	Standardized Estimate	Odds Ratio
INTERCPT	1	-0.2287	0.0212	115.8326	0.0001	N/A	0.796

Residual Chi-Square = 5471.3181 with 6 DF (p=0.0001)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
SMC	3.9802	0.0460
CMC	63.7890	0.0001
CNMC	337.2833	0.0001
WMC	1.3011	0.2540
SNP	4760.0395	0.0001
DIST	2370.1774	0.0001

Step 1. Variable SNP entered:

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald χ^2	Pr > χ^2	Standardized Estimate	Odds Ratio
INTERCPT	1	-3.9074	0.0821	2263.2626	0.0001	N/A	0.020
SNP	1	5.5818	0.1025	2963.6745	0.0001	1.249383	265.549

Association of Predicted Probabilities and Observed Responses

Concordant = 86.2%	Somers' D = 0.807
Discordant = 5.5%	Gamma = 0.880
Tied = 8.4%	Tau-a = 0.398
(19890000 pairs)	c = 0.903

Residual Chi-Square = 1307.8369 with 5 DF (p=0.0001)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
SMC	0.5463	0.4598
CMC	26.8515	0.0001
CNMC	57.2080	0.0001
WMC	0.0133	0.9081
DIST	1252.2815	0.0001

Step 2. Variable DIST entered:

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald χ^2	Pr > χ^2	Standardized Estimate	Odds Ratio
INTERCPT	1	-2.2290	0.0864	666.2145	0.0001	N/A	0.108
SNP	1	5.3776	0.1093	2420.5782	0.0001	1.203667	216.492
DIST	1	-3.7943	0.1205	990.8841	0.0001	-0.697531	0.022

Association of Predicted Probabilities and Observed Responses

Concordant = 94.4%	Somers' D = 0.891
Discordant = 5.3%	Gamma = 0.894
Tied = 0.3%	Tau-a = 0.440
(19890000 pairs)	c = 0.946

Residual Chi-Square = 69.2759 with 4 DF (p=0.0001)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
SMC	0.1606	0.6886
CMC1	0.3580	0.0013
CNMC	52.6975	0.0001
WMC	4.5614	0.0327

Step 3. Variable CNMC entered:

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald χ^2	Pr > χ^2	Standardized Estimate	Odds Ratio
INTERCPT	1	-2.9557	0.1360	472.3933	0.0001	N/A	0.052
CNMC	1	0.8779	0.1215	52.2365	0.0001	0.161813	2.406
SNP	1	5.3016	0.1092	2357.1294	0.0001	1.186663	200.655
DIST	1	-3.8112	0.1214	985.5028	0.0001	-0.700643	0.022

Association of Predicted Probabilities and Observed Responses

Concordant = 94.5%	Somers' D = 0.893
Discordant = 5.2%	Gamma = 0.895
Tied = 0.3%	Tau-a = 0.441
(19890000 pairs)	c = 0.946

Residual Chi-Square = 16.5744 with 3 DF (p=0.0009)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
SMC	0.1708	0.6794
CMC	11.6400	0.0006
WMC	4.7056	0.0301

Step 4. Variable CMC entered:

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald χ^2	Pr > χ^2	Standardized Estimate	Odds Ratio
INTERCPT	1	-4.1674	0.3807	119.8055	0.0001	N/A	0.015
CMC	1	1.2086	0.3533	11.7014	0.0006	0.087071	3.349
CNMC	1	0.8896	0.1215	53.5804	0.0001	0.163974	2.434
SNP	1	5.3009	0.1093	2352.3886	0.0001	1.186515	200.523
DIST	1	-3.7935	0.1216	973.6082	0.0001	-0.697395	0.023

Association of Predicted Probabilities and Observed Responses

Concordant = 94.6%	Somers' D = 0.894
Discordant = 5.2%	Gamma = 0.896
Tied = 0.2%	Tau-a = 0.441
(19890000 pairs)	c = 0.947

Residual Chi-Square = 5.0366 with 2 DF (p=0.0806)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
SMC	0.0524	0.8189
WMC	4.9447	0.0262

Step 5. Variable WMC entered:

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald χ^2	Pr > χ^2	Standardized Estimate	Odds Ratio
INTERCPT	1	-3.4248	0.5050	45.9982	0.0001	N/A	0.033
CMC	1	1.2221	0.3544	11.8886	0.0006	0.088044	3.394
CNMC	1	0.8912	0.1215	53.7689	0.0001	0.164272	2.438
WMC	1	-0.7600	0.3408	4.9727	0.0257	-0.045684	0.468
SNP	1	5.3007	0.1093	2351.8317	0.0001	1.186475	200.487
DIST	1	-3.8081	0.1219	976.5818	0.0001	-0.700064	0.022

Association of Predicted Probabilities and Observed Responses

Concordant = 94.5%	Somers' D = 0.893
Discordant = 5.2%	Gamma = 0.896
Tied = 0.3%	Tau-a = 0.441
(19890000 pairs)	c = 0.947

Residual Chi-Square = 0.0794 with 1 DF (p=0.7782)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
SMC	0.0794	0.7782

NOTE: No additional variables met the 0.05 significance level for entry into the model.

AD-33 Bookplate
(1-63)

U. S. FOREST SERVICE

NATIONAL

FT. COLLINS

AGRICULTURAL

LIBRARY

LIBRARY COPY

ROCKY MT. FOREST & RANGE

EXPERIMENT STATION